AD_____

Award Number: W81XWH-08-1-0383

TITLE: A Genome-wide Breast Cancer Scan in African Americans

PRINCIPAL INVESTIGATOR: Christopher A. Haiman, ScD

CONTRACTING ORGANIZATION: University of Southern California
Los Angeles, CA 90033

REPORT DATE: June 2009

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01/05/2009	Annual	1 June 2008- - 31 May 2009

4. TITLE AND SUBTITLE
A Genome-wide Breast Cancer Scan in African Americans

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-08-1-0383

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Christopher A. Haiman

email: haiman@usc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Southern
California
Los Angeles, California
90033

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Material Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This genome-wide association of breast cancer includes over 3,000 African American women with invasive disease and over 3,000 age-matched African American controls from many existing case-control studies of breast cancer in the U.S. In the first year of this project all DNA samples from these studies were sent to Dr. Haiman's laboratory at the University of Southern California, quantitated and arrayed for genotyping. We have also assembled the covariate file that contains established breast cancer risk factor data collected from these studies, and the data have been standardized for the analysis. As of April 1, 2009, we have genome-wide scanned 2,200 cases and 2,200 controls from a number of the participating studies using the Illumina 1M Beadchip. We have assessed the blinded quality control replicates and overall call rates by SNP and sample and the data appear to be of very high quality. We have conducted a very preliminary statistical analysis of the data and there are a number of promising signals including what we believe is a novel risk locus for breast cancer that may be of particular importance for estrogen receptor negative breast cancer. The analysis will continue as additional samples/studies are scanned over the next year.

15. SUBJECT TERMS
Breast cancer, estrogen receptor negative, genome-wide association study

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	16	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

Introduction

Genome-wide association studies of breast cancer have been completed or are in progress among populations of European ancestry, and several regions have been identified that appear to contribute susceptibility to this cancer. Recent data suggests that not all risk alleles for common cancers will be revealed however by studies limited to Whites of European ancestry, and that similar efforts in other racial and ethnic populations will be needed to identify the full spectrum of common risk alleles that contribute to disease risk in the population. We propose to identify genetic risk alleles for breast cancer risk among African American women by performing a well-powered whole-genome association scan. For this project we have established a collaborative network of investigators whose careers have been dedicated to studying breast cancer in minority populations who have contributed samples and covariates from each of their respective studies to identify genetic variants that contribute to risk of breast cancer in this minority population. As described below, we are in the progress of genotyping and analyzing 1.1 million SNPs examined in 3,000 African American breast cancer cases and 3,000 controls.

BODY

The Specific Aim of this application is to identify genetic risk alleles for breast cancer among African American women by performing a well-powered genome-wide association study (GWAS). For this project, I have established a network of leaders in the breast cancer research community with long-standing interests in breast cancer research in African Americans, all of whom have existing case-control studies of breast cancer in the U.S. A list of the collaborators who are participating in this study is provided in Table 1. Funding for the genotyping of samples from the MEC, CARE, WCHS, SFBC and BCFR studies is covered by this DOD-BCRP grant. The genotyping of the other studies has been provided by a number of other sources. All together these studies include over 4,600 African American women with invasive breast cancer and over 4,500 age-matched African American controls with available DNA.

Table 1. African American Breast Cancer Cases From Participating Studies.

Studies:	**PI**	**Cases/Controls**
Multiethnic Cohort (MEC)	L. Kolonel	686/996
The Women's Contraceptive and Reproductive Experiences Study (CARE-Los Angeles)	L. Bernstein	347/217
The Women's Circle of Health Study (WCHS)	C. Ambrosone	272/240
The San Francisco Bay Area Breast Cancer Study (SFBC) / Northern California component of the Breast Cancer Family Registry (BCFR)	E. John	607/279
The Carolina Breast Cancer Study (CBCS)	R. Millikan	656/608
The Nashville Breast Health Study (NBHS)	W. Zheng	310/186
Prostate, Lung, Colon, Ovarian Screening Trial (PLCO)	R. Hoover	72/144
Wake Forest University Breast Cancer Study (WFUBCS)	J. Hu	125/153
Black Women's Health Study (BWHS)	J. Palmer	667/825
Women's Insights and Shared Experiences Study (WISE)	T. Rebbeck	145/367
The Nigerian Breast Cancer Study	F. Olopade	797/491
	Total:	**4,684/4,506**

Genotyping Quality Control

To date, in stage 1 of this scan, we have genotyped 2,280 cases and 2,316 controls from the MEC, CARE, WCHS, SFBC, BCFR and CBCS studies using the Illumina 1M array. We have observed SNP call rates >95% for 96.3% of samples (samples with call rates <95% were removed from the preliminary analysis discussed below). The average concordance rates of blinded duplicates (n=107; 2%) was 99.96% (all pairs were >99.35%). Of the 1,155,397 SNPs on the array we made the following exclusions:

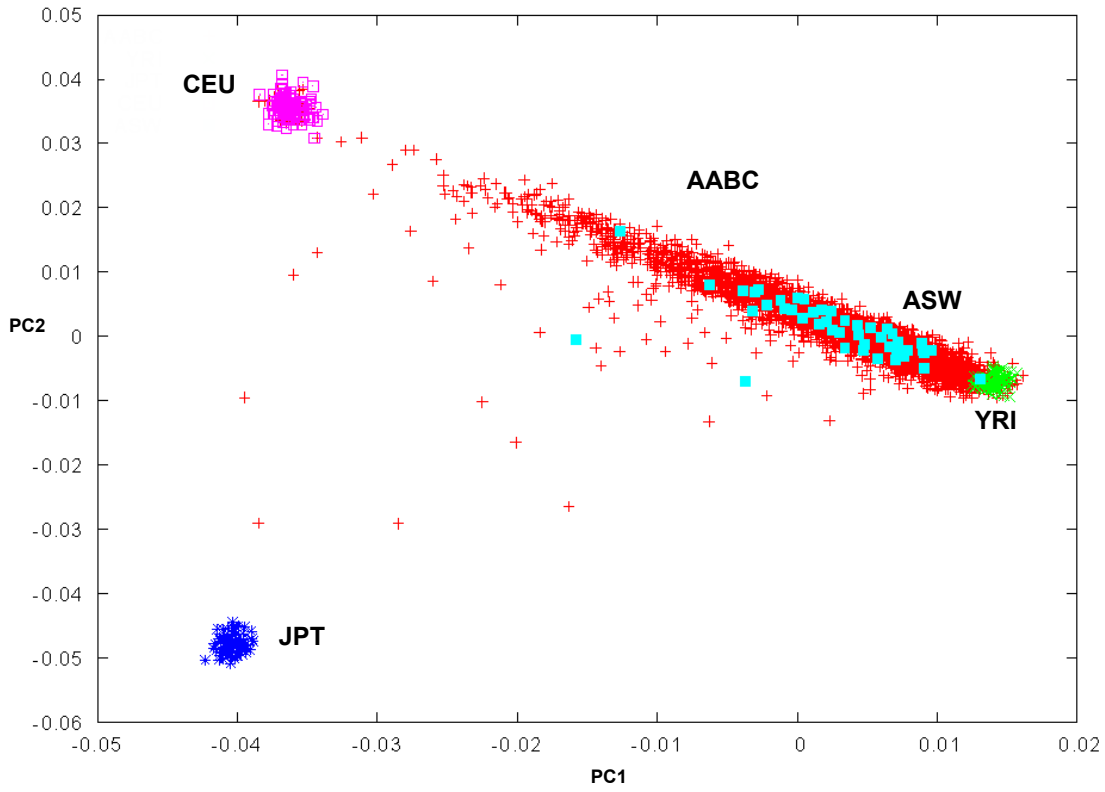
1) 54 provided no genotypes (most likely due to a technical failure of the assay)
2) 2,744 were monomorphic (0.2%)
3) 18,744 had call rates <95% (1.6%), which is a standard threshold for exclusion
4) 1,565 with <98% concordance based on the blinded duplicates (0.1%)

5) 186,649 with minor allele frequencies <5% (~15%). These less common SNPs will be evaluated once we have completed the scanning of all samples in stage 1.
After these exclusions 946,439 SNPs with frequencies >5% remained for analysis among the 2,197 cases and 2,230 controls.

Assessment of Population Structure

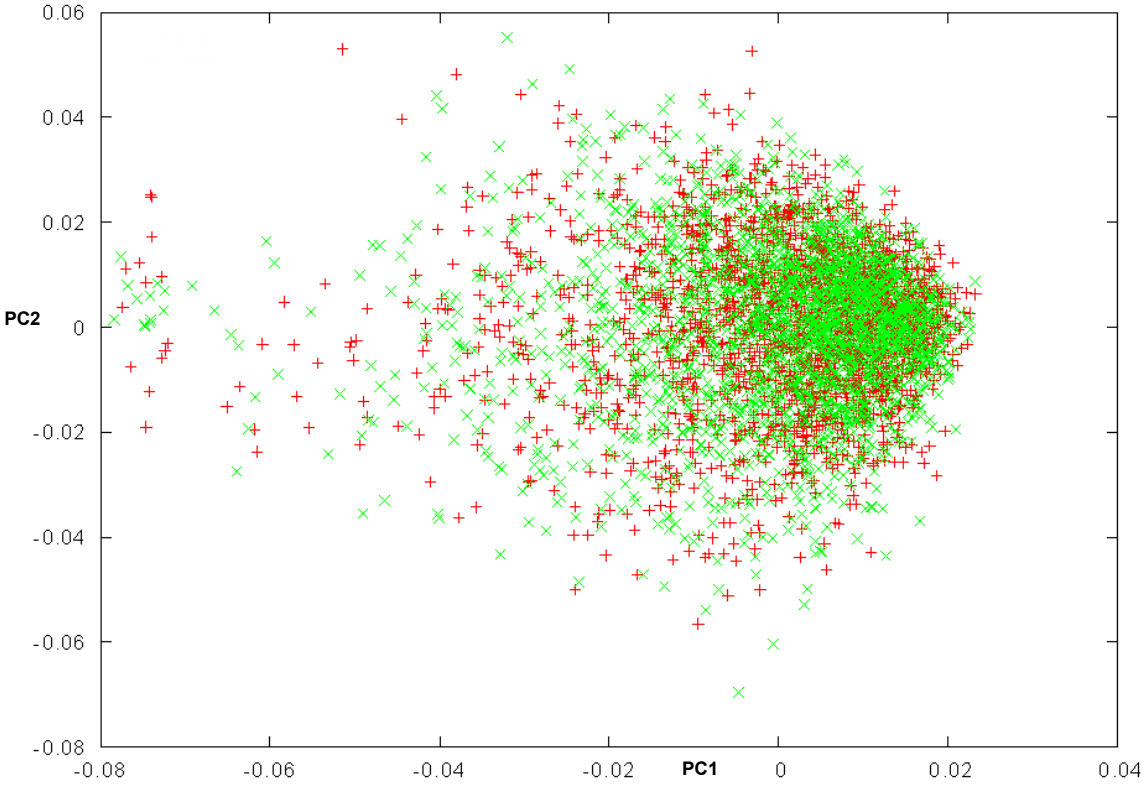
In the analysis, we began by examining the potential confounding effects of population stratification. To do so, we selected 10,000 SNPs at random from the >1 million SNPs on the array and utilized principle components analysis (1) to evaluate the genetic background of each subject. Figure 1 is a plot of the 1st versus the 2nd principle components from this analysis. For reference, we have provided the phase 2 HapMap populations (2): the CEUs of European ancestry (pink); the YRIs (Yoruban) population from West Africa (green), and; the Japanese sample (JPT, blue). Also illustrated is the African American sample from phase 3 HapMap (ASW, light blue) and the ~4,400 African Americans subjects from our GWAS (AABC, red). As shown below, the 3 phase 2 HapMap populations can be clearly distinguished based on these 1st 2 principle components. Both of the African American samples show a similar pattern, with the majority of the sample clustering near the YRIs, as expected, with a trail towards the CEU cluster. The trail is an indication of the degree of European admixture in the African American population, which is well-captured by the 1st principle component.

Figure 1. Principal Components Analysis (PCA):
10K Random SNPs: AABC GWAS + HapMap Panels



Next, we examined the degree to which population structure differs between the African American breast cancer cases and African American controls in our study. We conducted principle components analysis limited to these ~4,400 subjects and evaluated all pair-wise comparisons between the first 10 principle components. Figure 2 is a plot of the 1st versus the 2nd principle components for these subjects, with the cases in red and the controls in green. Here we observed quite good overlap between cases and controls, which suggests that there are unlikely large differences in population structure between the groups. This pattern was also observed for all pair-wise comparisons of the 1st 10 eigenvectors. The 1st principle component was by far the most important component in explaining genetic differences between subjects with an eigenvalue >70 (the 2nd principle component had an eigenvalue of ~5). The test for a case-control difference for the 1st principle component was non-significant (p=0.10), which suggests that the degree of admixture in cases is similar to that in controls. Tests of case-control differences for the higher order eigenvectors were are not statistically significant. In summary, based on this preliminary analysis, population stratification does not appear to be a critical issue in this particular study.

Figure 2. Principal Components Analysis (PCA): Differences by Case-Control Status



Association Testing

In the statistical analysis we conducted a 1-df test of trend for the ~950,000 common SNPs and compared the observed versus the expected distribution of the test statistics. In the unadjusted analysis, we observed a slight over-dispersion of p-values, with a lambda inflation factor of 1.04. However, correcting for study, age and the 1st principal component reduced the lamba to 1.01, with there being virtually no remaining over-dispersion. The quantile-quantile plot for the adjusted analysis is provided in Figure 3. This initial analysis includes approximately two-thirds of the total number of subjects that will be scanned (target = 3,000 cases and 3,000 controls) and so far we see no excess of very small p-values. We did note a very slight excess of associations at p<0.05 than what would be expected based on chance, especially in the 10^{-3} to 10^{-5} range (Table 2).

Figure 3. Quantile-Quantile Plot

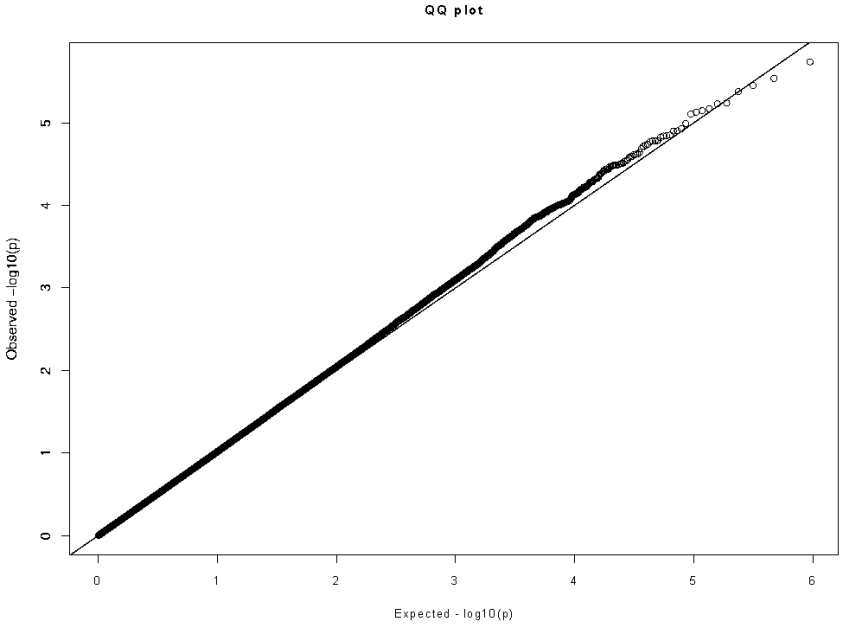


Table 2. Observed vs. Expected Number of Nominally Significant Associations.

Level of significance	O	E	O/E
0.01-0.05	39,353	37,858	1.04
0.001-0.01	9,262	8,518	1.09
0.0001-0.001	1,045	851	1.23
0.00001-0.0001	121	85	1.42
<0.00001	10	10	1.00
P <0.05	49,791	47,322	1.05

In the preliminary analysis, 5 SNPs had p-values of 10^{-5} - 10^{-6} (Table 3), with relative risks per allele ranging from 1.21-1.25. We attempted to replicate these associations in an additional 812 African American breast cancer cases and 1,192 controls from the BWHS and WISE studies (listed in Table 1). As shown in Table 3, none of these SNPs replicated with p<0.05. However, we did note two borderline significant associations with SNP2 (p=0.06) and SNP5 (p=0.07). Interestingly, these two SNPs were correlated and located in very close proximity in the same region of the genome. Based on these findings we took a closer look at this 200kb region and observed there to be 14 SNPs with p-values $<10^{-4}$ in the region. All of these SNPs lie in the same block of linkage disequilibrium in the middle of interesting candidate gene. We are now in the process of resequencing the coding regions of this gene in search of a functional coding variant.

Table 3. Replication Testing of the 5 Most Significant Allelic Associations.

		Stage 1 Sample 2,197 cases and 2,230 controls		Replication Sample 812 cases and 1,192 controls	
Risk Variant	Risk Allele Frequency	OR(95% CI)	p-value(1df)	OR(95% CI)	p-value(1df)
SNP1	0.65	1.24(1.13-1.36)	2.73x10-6	0.93(0.81-1.06)	0.28
SNP2	0.70	1.25(1.13-1.38)	7.12x10-6	1.14(0.99-1.31)	0.06
SNP3	0.42	1.21(1.11-1.32)	1.87x10-5	1.02(0.89-1.16)	0.78
SNP4	0.35	1.21(1.10-1.33)	5.12x10-5	1.06(0.93-1.21)	0.40
SNP5	0.68	1.21(1.10-1.32)	7.50x10-5	1.13(0.99-1.30)	0.07

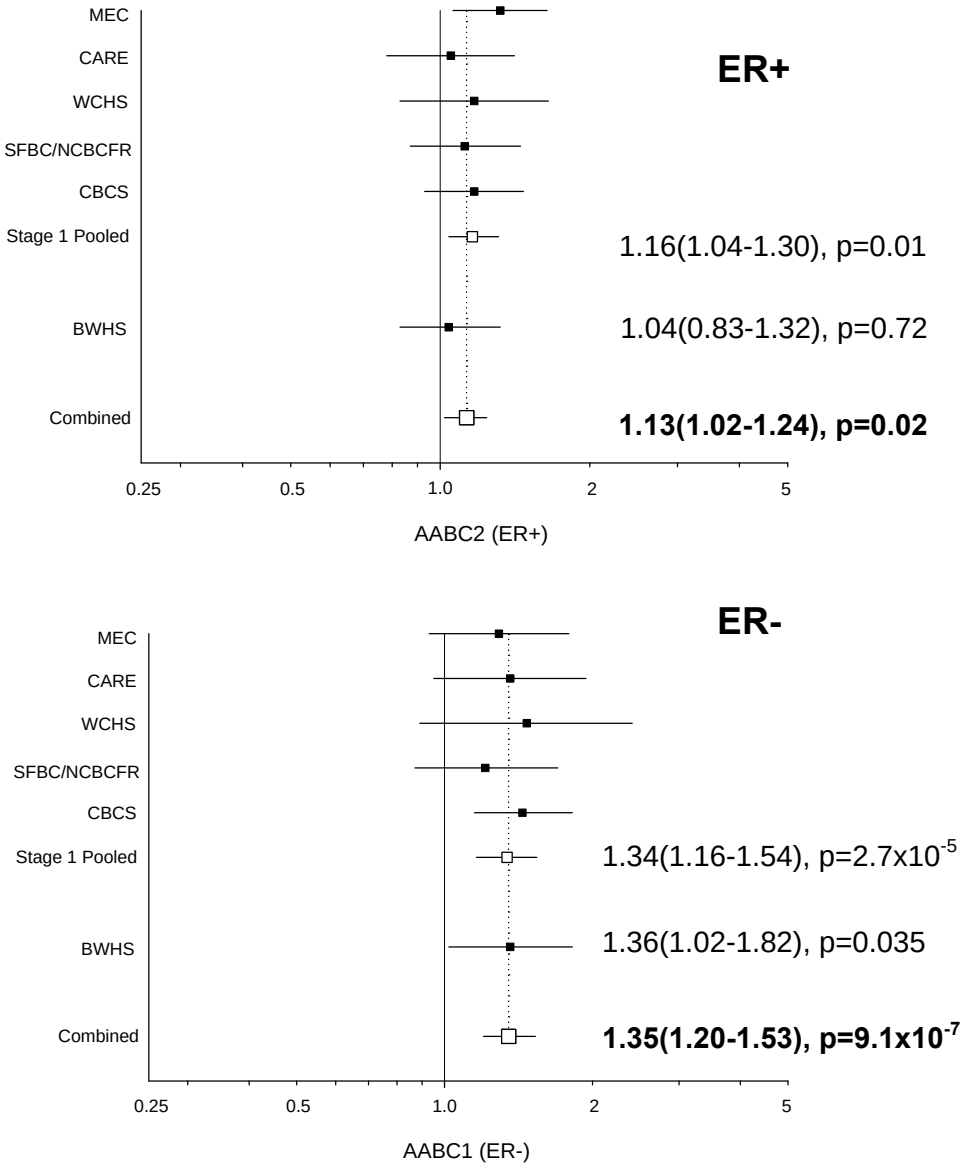
To better understand the associations at this novel breast cancer susceptibility locus we conducted a number of additional analyses. We began with stratifying the analysis by estrogen receptor status, as many of the risk variants identified in previous GWAS in Whites have stronger associations with estrogen receptor positive tumors (3). This analysis included 750 cases with estrogen receptor negative disease and 1,070 cases with estrogen receptor positive disease. For the ~50 markers on the Illumina 1M array that we examined in this region we observed no significant evidence of heterogeneity by estrogen receptor status. However, we did detect a much stronger signal in the region for estrogen receptor negative tumors with there being 21 SNPs with p-values less than 10^{-4} compared to only 3 at this level of significance for estrogen receptor positive disease. Based on these findings we went back to the replication sample and stratified the analysis by estrogen receptor status. In Figure 4, we provide the forest plots for SNP2 in a stratified analysis of the stage 1 sample and the replication sample (the top is for estrogen receptor positive breast cancer and the bottom is for estrogen receptor negative breast cancer). In the replication sample we had data for an additional 200 estrogen receptor positive cases and the association was non-significant, with the overall estimate of the stage 1 plus replication sample being marginally significant (OR, 1.13; 95% CI, 1.02-1.24; p=0.02). For estrogen receptor negative disease there were 140 cases in the replication sample and the association in this sample was nominally significant (OR, 1.36; 95% CI, 1.02-1.82; p=0.035) with the effect estimate nearly

identical to that observed in stage 1 (OR, 1.34; 95% CI, 1.16-1.54; $p=2.7x10^{-5}$). In the combined analysis of stage 1 plus the replication sample, which included ~900 African American breast cancer cases with estrogen receptor negative disease and >3000 African American controls, the effect was 1.36 with a p-value of $9.1x10^{-7}$. The test for heterogeneity of effect by estrogen receptor status was also significant at 0.01. *Although preliminary, these findings provide strong support for this region as harboring a risk variant that is particularly important for estrogen receptor negative breast cancer.* We are in the process of confirming these findings in the other African American studies as well as in other populations. Interestingly, the risk allele for this SNP has a frequency of 70% in African Americans and 30% in Whites. If confirmed, this locus may contribute to the higher proportion of estrogen receptor negative tumors in the African American population.

Fostering Additional Collaborations

Over the past year I have reached out to other epidemiologists with ongoing or planned studies of breast cancer in populations of African ancestry. This effort was recently acknowledged by NCI and in December of 2008 I was asked to participate in a meeting to discuss further collaborative efforts to study breast cancer among women of African ancestry. As a result, a number of additional investigators with samples from women in Nigeria, and samples from women of African ancestry in the Caribbean and elsewhere in the U.S. have joined our collaborative group. I am now leading this consortia effort. This is currently the largest and most definitive genetic study of breast cancer in African women ever conducted, with the increase in sample size allowing for even greater statistical power to detect alleles that contribute modest risk. Establishing a network of investigators interested in studying breast cancer in African American women was a crucial first step, and I am hopeful that it will lead to many additional collaborative studies that will be successful in defining the role of genes and environment in the etiology of breast cancer in women of African ancestry.

Figure 4. The Association of SNP2 by Estrogen Receptor Status.



Key Research Accomplishments

- Establishing an NCI-recognized consortia to study breast cancer among women of African ancestry
- Fostering additional collaborations with breast cancer researchers outside of those listed in my DOD application who are participating in the GWAS
- Conducting the first genome-wide association study of breast cancer among African American women
- Identifying the first genetic risk locus for estrogen receptor negative breast cancer

Reportable Outcomes

- The findings discuss in this progress report were presented at the American Association of Cancer Research annual conference in Denver, CO (April, 2009)

Conclusion

This is the first genetic risk factor for estrogen receptor negative breast cancer. These findings clearly highlight the importance of conducting GWAS of breast cancer in other racial/ethnic populations, particularly African Americans and Latinos, who are more likely to be diagnosed with estrogen negative disease.

A similar collaborative effort to conduct a GWAS of breast cancer among Latinas is also needed. This is a crucial step towards understanding the disproportionate rate of estrogen receptor negative disease in this understudied minority population. I have recently applied to the DOD-BCRP EOH program to conduct such a study.

Revealing the genetic causes of breast cancer *in each population*, and more importantly, inherited susceptibility for aggressive disease subtypes, will in time translate into more targeted preventive measures and treatment strategies for those not only at risk of developing the disease but also for those most likely to die from the disease.

References

1. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38(8):904-9.
2. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. Nature 2005;437(7063):1299-320.
3. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet 2008;4:e1000054.

Appendices

NA